

OEG Publication

Ramos JA

Descubrimiento automático de mappings

I Simposio de Doctorado en Web Semántica
CAEPIA 2007: Conferencia de la Asociación Española para la Inteligencia Artificial
Díaz Rubio y Gómez Berbís eds. Edited by Universidad de Salamanca.
November 13th, 2007.
Salamanca, Spain.
ISBN: 978-84-611-8856-7

Descubrimiento automático de mappings

José Ángel Ramos Gargantilla

OEG - Facultad de Informática. Universidad Politécnica de Madrid (UPM)
Campus de Montegancedo, s/n. 28660 Boadilla del Monte. Madrid. Spain
{jarg@fi.upm.es}

Resumen. Dentro de la problemática de la integración de información, los elementos claves son los mappings, unidades que relacionan las diferentes representaciones (ontologías, bases de datos, redes semánticas, etc.). Y dentro de toda la colección de operaciones que los mappings llevan asociadas en todo su ciclo de vida, el cuello de botella se encuentra en su descubrimiento. Con este trabajo doctoral se pretende dar un paso más en este campo realizando un nuevo modelo de mappings lo menos limitado, y a la vez funcional, posible a diferentes representaciones y lo más versátil para la combinación de técnicas de descubrimiento, de toda índole, ya existentes y de nuevo cuño de manera automática, basándose en un sistema experto previamente construido a costa de evaluaciones sobre casos de uso reales.

1. Antecedentes y estado actual

Las ontologías son conceptualizaciones que permiten capturar conocimiento consensuado de un dominio dado de una forma genérica y formal, para ser reutilizado y compartido por aplicaciones y grupos de personas [38]. De esta definición se puede inferir erróneamente que hay una sola ontología para modelar un dominio. Sin embargo, se pueden encontrar en la literatura varias ontologías que modelan el mismo conocimiento de diferentes formas. Así, en el campo del comercio electrónico hay varios estándares e iniciativas conjuntas para la clasificación de productos y servicios (UNSPSC ¹, e-cl@ss ², RosettaNet ³, etc.). Esta heterogeneidad en ontologías también se presenta en otros muchos dominios [16]: lingüística (WordNet, EuroWordNet, GUM, Mikrokosmos, SENSUS), medicina (GALEN, UMLS, ON9), química (Chemicals, Ions), etc.

Para resolver esta heterogeneidad se recurre a multitud de modelos y métodos para integración, mezcla, articulación, etc. Todos estos procesos llevan más o menos implícito el uso de mappings. Los mappings son entidades de conocimiento que relacionan componentes de dos conceptualizaciones mediante una función. Serán los mappings el medio de expresar la igualdad entre conceptos de dos ontologías, la descomposición de un concepto en otros de otra conceptualización, la equivalencia

¹ <http://www.unspc.org>

² <http://www.eclass.de>

³ <http://www.rosettanet.org>

matemática de unos atributos y otros, etc. Casi el único requisito establecido para la realización de mappings entre fuentes de conocimiento es que sus dominios estén solapados en alguna medida.

Los mappings son la base de la interoperabilidad e integración de fuentes de conocimiento. Debido al gran ámbito en el que pueden aparecer, los mappings se han visto tradicionalmente desde muchos puntos de vista. Podemos encontrar definiciones de mapping de manera genérica sobre estructuras [34] y, más concretamente, sobre ontologías desde diferentes enfoques: similitud semántica [39], transformación de instancias [23], métodos de resolución de problemas [7], alineamiento de ontologías [4], morfismos entre ontologías [19], etc.

La gran versatilidad de los mappings viene avalada por la cantidad de procesos en los que son parte fundamental (obviemos por ahora en todos aquellos en los que es una fuente de conocimiento más). La literatura está plagada de sistemas, modelos y métodos en los que intervienen directamente los mappings: *Ontology Alignment* [32] [37] [11], *Ontology Mapping* [9] [31], *Ontology Matching* [8] [1], *Ontology Integration* [26] [37] [35], *Ontology Merging* [33] [35] [9] [17], *Ontology Articulation* [29], *Ontology Coordination* [36], *Ontology Translation* y *Transformation* [9] y *Ontology Reconciliation* [4].

Y todo esto sin salirnos del ámbito ontológico. Si tomamos la definición de mapping general para cualquier representación del conocimiento y cualquier número de las mismas, el número de escenarios posibles se dispara exponencialmente.

Como se puede ver, la utilidad de los mappings está más que demostrada y su gran versatilidad y proyección es indiscutible. Sin embargo, el gran problema de los mappings es su creación. El descubrimiento automatizado del conocimiento que es representado por esos mappings y su modelización es una de las líneas de investigación abiertas. Los trabajos que hasta ahora se han realizado están enfocados a resolver fundamentalmente la búsqueda de mappings para realizar uno de los procesos antes citados (*Integration*, *Merging*, *Articulation*, etc.). Tan sólo el proceso de *Ontology Alignment*, resuelto en muchas ocasiones por métodos de *Ontology Matching*, se enfoca al descubrimiento de mappings no dirigidos por una finalidad.

Existen muchos trabajos y algoritmos que realizan *Ontology Matching* (*TranScm* [2r8], *H-Match* [6], *Anchor-Prompt* [33], *Cupid* [22], *S-Match* [15], *T-tree* [10], *OLA* [14], *NOM & QOM* [13] [12], *Falcon-AO* [18], *DSSim* [41], *COMA++* [25], *AUTOMS* [20], *JHU/APL* [3], *PRIOR* [24], *RiMOM* [21], *OWL-CtxMatch* [30], *NLM* [40], etc.), algunos de ellos con resultados muy buenos para determinados ámbitos u objetivos.

A pesar de lo que se ha trabajado en diferentes métodos, queda todavía mucho por hacer, sobre todo en el descubrimiento de mappings con semántica más allá de la similitud, la disyunción y la subsunción.

2. Objetivos concretos de la investigación

Es en el descubrimiento automático de mappings donde se enmarca el presente trabajo. Para ello, se pretenden alcanzar diferentes objetivos:

Objetivos metodológicos:

- **M1:** Modelización y formalización de los mappings.
- **M2:** Desarrollo de nuevos métodos y técnicas de descubrimiento automático de mappings.
- **M3:** Desarrollo de un sistema experto de combinación de los métodos y técnicas disponibles.

Objetivos tecnológicos:

- **T1:** Diseño de una arquitectura software que integre los recursos y métodos del trabajo de investigación.
- **T2:** Implementación de los métodos y técnicas desarrollados en el objetivo M2.
- **T3:** Integración de herramientas y recursos en la arquitectura.
- **T4:** Evaluación del sistema de descubrimiento.

Para lograr el objetivo M1 se está realizando un profundo estudio de la cuestión sobre mappings y un estudio pormenorizado de las actuales representaciones (tanto a través de lenguajes propiamente dichos como a través de instancias de modelos de mappings, es decir, desde OWL-C [5] hasta SKOS-Mapping [27] pasando por R₂O [2]). De todo este conocimiento, obtenido a través de artículos de investigación, tesis doctorales y los propios casos de uso, se obtendrá la modelización y formalización de los mappings, unida al lenguaje (probablemente de nueva acuñación) para la expresión de los mappings siguiendo la modelización acordada.

Dentro de las tareas para cumplir el objetivo M2 se ha realizado un estudio detallado de las técnicas y métodos existentes para el descubrimiento de mappings. Con este estudio finalizado, la siguiente tarea es la determinación de los ámbitos no cubiertos y de las características mejorables dentro del descubrimiento. Para esto se parte de iniciativas como Ontology Alignment Evaluation Initiative (OAEI⁴). Se procederá a plantear y desarrollar métodos y técnicas que cubran estas lagunas, dirigidos a casos de uso reales para contar con el apoyo de una experimentación real que pueda aportar la retroalimentación conveniente y necesaria en todo este tipo de técnicas y heurísticas. Por el momento, se está trabajando en un caso de uso real, mediante un convenio de colaboración con el Instituto Geográfico Nacional, para la creación de un sistema integrado de una ontología y diferentes bases de datos en las que aparecen parámetros sobre los que hacer el descubrimiento de análisis léxico, sintáctico (expresiones regulares, lematizado), semántico (hiperonimia, sinonimia), estructural, información propia de los atributos concretos para el caso de uso, etc.

El objetivo T1 se alcanzará tomando como base los resultados que componen el objetivo M1 y M2. Basándose en el estudio de los mappings y sus características se diseñará una arquitectura software que integre los diferentes recursos necesarios (ontologías, tesauros, otras conceptualizaciones, documentos, etc.) y en la que haya cabida para los métodos que actúan sobre esos recursos. Es el diseño de la arquitectura y las interacciones de sus módulos donde residirá gran parte del éxito de la tecnología, puesto que será crucial para que la lógica de negocio tenga las

⁴ <http://oaei.ontologymatching.org/>

facilidades necesarias de combinatoria de métodos y técnicas para los casos concretos que se le presenten.

En un primer boceto del sistema (Figura 1), se exponen los componentes funcionales principales, quedando las dos actividades del sistema imbricadas a través de los componentes compartidos: descubrimiento de mappings y evaluación de los resultados obtenidos, bien de manera también automática (por medio de conjuntos resueltos previamente) o bien de manera manual a cargo de un experto en el dominio. Así, residirá en el módulo Descubrimiento el sistema experto (objetivo M3) que resolverá, a partir de los recursos de entrada y los métodos, técnicas y recursos de apoyo disponibles en cada momento (los componentes Técnicas de descubrimiento y Entrada/Salida de las Fuentes), cuál será la secuencia de ejecución y la consiguiente combinación de resultados para la obtención de un resultado conjunto y óptimo.

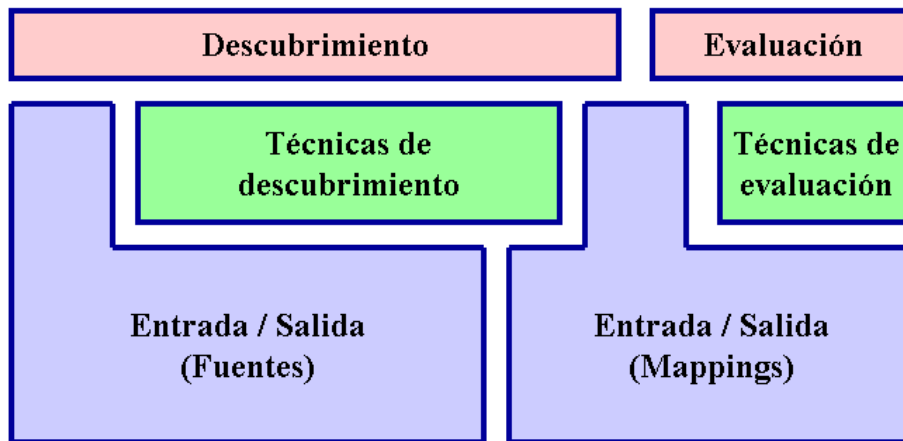


Figura 1. Componentes del sistema

El componente de descubrimiento tendrá a su disposición diferentes gestores (tanto de fuentes para mapear como de fuentes de apoyo a las técnicas), las técnicas de descubrimiento y el gestor de Entrada/Salida de los Mappings (siguiendo el modelo y lenguaje determinado con el objetivo M1).

El componente de evaluación tendrá acceso a los gestores de Entrada/Salida de Mappings (siempre sujeto al modelo y lenguaje fruto del objetivo M1), a las técnicas de evaluación y además podrá devolver al usuario del sistema el resultado de la evaluación.

Dentro de la arquitectura fruto de las tareas del objetivo T1, se procederá a la implementación de los métodos y técnicas de descubrimiento descritos en el objetivo M2, consiguiendo así el objetivo T2.

Para alcanzar el objetivo T3 se realizarán las siguientes tareas: integración en la arquitectura del objetivo T1 de los recursos disponibles (ontologías, documentos, etc.) y de las herramientas, tanto de apoyo para el descubrimiento como de descubrimiento en sí. Debido a la existencia de métodos y técnicas de descubrimiento con apoyo lingüístico, aparecerán integrados recursos como vocabularios, diccionarios, lexicones o EuroWordnet; anotadores de texto como DataLexica o Connexor FDG; y procesadores de anotaciones como OntoTagger u ODELinger, éstos últimos

desarrollados en el Grupo de investigación con financiación nacional (TIC-2001-2745) y europea (IST-2001-34373). Finalmente se evaluarán los métodos y técnicas de descubrimiento en el objetivo T4 para la evaluación propia del descubrimiento.

3. Conclusiones

Al encontrarse el trabajo de tesis en sus primeras fases, de documentación y perfilamiento de objetivos y alcance, son pocas, y preliminares, las conclusiones del trabajo realizado hasta el momento.

Se ha definido un objetivo genérico y se ha desglosado en diversos objetivos parciales metodológicos y tecnológicos a los que todavía les falta por definir de manera precisa su alcance. De estos objetivos, ya se ha trazado una primera versión de lo que sería la relación de los grandes módulos funcionales que formarán la tecnología. Se está trabajando también en la modelización y gestión de los mappings que el sistema descubrirá y evaluará.

Además, se están identificando y definiendo casos de uso reales para la evaluación tanto de los métodos como de la tecnología que se desarrolle en entornos fuera de los prototipos y ejemplos de investigación.

4. Referencias

- [1] Avesani P, Giunchiglia F, Yatskevich M (2005) A Large Scale Taxonomy Mapping Evaluation. In Proceedings of International Semantic Web Conference (ISWC), 2005.
- [2] Barrasa J, Corcho O, Gómez-Pérez A (2004) R2O, an extensible and semantically based database-to-ontology mapping language. In Chris Bussler and Val Tannen, editors, Proceedings of the 2nd Workshop on Semantic Web and Databases, Toronto, Canada, August 2004.
- [3] Bethée W, Fink C, Beecher-Deighan J (2006) JHU/APL Onto-Mapology Results for OAEI 2006. Results of the 2006 Ontology Alignment Contest. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [4] Bouquet P (coordinator), Euzenat J, Franconi E, Serafini L, Stamou G, Tessaris S (2004) D2.2.1 Specification of a common framework for characterizing alignment Knowledge Web (FP6-507482). <http://www.knowledgeweb.net/>
- [5] Bouquet P, Giunchiglia F, van Harmelen F, Serafini L, Stuckenschmidt H (2003) C-OWL: Contextualizing ontologies. In D. Fensel, K. Sycara, and J. Mylopoulos, editors, Proceedings of the Second International Semantic Web Conference, number 2870 in Lecture Notes in Computer Science, pages 164–179. Springer Verlag, October 2003.
- [6] Castano S, Ferrara A, Messa G (2006) Results of the HMatch Ontology Matchmaker in OAEI 2006. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [7] Crubézy M, Musen MA (2003) Ontologies in Support of Problem Solving. Handbook on Ontologies. S. Staab and R. Studer, Springer: 321-341.
- [8] Doan A, Madhavan J, Domingos P, Halevy A (2002) Learning to Map between Ontologies on the Semantic Web. In WWW '02: Proceedings of the 11th international conference on World Wide Web, pages 662–673, New York, NY, USA, 2002. ACM Press.

- [9] Dou D, Drew M, Qi P (2003) Ontology Translation on the Semantic Web. In Proceedings of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE 2003). LNCS 2888. Springer-Verlag. Catania (Sicily), Italy. 2003. Pgs. 952-969.
- [10] Euzenat J (1994) Brief overview of T-tree: the Tropes taxonomy building tool. In Proc. 4th ASIS SIG/CR Workshop on Classification Research, pages 69–87, Columbus (OH US), 1994.
- [11] Euzenat J (2004) An API for Ontology Alignment. In Proc. 3rd international semantic web conference, Hiroshima (JP), pages 698–712, 2004.
- [12] Ehrig M, Staab S (2004) QOM – quick ontology mapping. In Proc. 3rd International Semantic Web Conference (ISWC), volume 3298 of Lecture notes in computer science, pages 683–697, Hiroshima (JP), 2004.
- [13] Ehrig M, Sure Y (2004) Ontology mapping – an integrated approach. In Proc. 1st European Semantic Web Symposium (ESWS), volume 3053 of Lecture notes in computer science, pages 76–91, Hersounisous (GR), May 2004.
- [14] Euzenat J, Valtchev P (2004) Similarity-based ontology alignment in OWL-lite. In Proc. 15th European Conference on Artificial Intelligence (ECAI), pages 333–337, Valencia (ES), 2004.
- [15] Giunchiglia F, Shvaiko P (2003) Semantic matching. The Knowledge Engineering Review (KER), 18(3):265–280, 2003.
- [16] Gómez-Pérez A, Fernández-López M, Corcho O (2003) Ontological Engineering, Springer-Verlag, 2003
- [17] Hitzler P, Krötzsch M, Ehrig M, Sure Y (2004) What Is Ontology Merging? In proceedings of the First International Workshop on Contexts and Ontologies: Theory, Practice and Applications (C&O). Workshop at the 20th National Conference on Artificial Intelligence, AAAI-05, Pittsburgh, Pennsylvania, July 2005, pp. 104-107.
- [18] Hu W, Cheng G, Zheng D, Zhong X, Qu Y (2006) The Results of Falcon-AO in the OAEI 2006 Campaign. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [19] Kalfoglou Y, Schorlemmer M (2005) Ontology Mapping: The State of the Art. Dagstuhl Seminar on Semantic Interoperability and Integration 2005, Germany, 2005.
- [20] Kotis K, Valarakos A, Vouros G (2006) AUTOMS: Automated Ontology Mapping through Synthesis of Methods. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [21] Li Y, Li J, Zhang D, Tang J (2006) Result of Ontology Alignment with RiMOM at OAEI'06. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [22] Madhavan J, Bernstein P, Rahm E (2001) Generic schema matching using Cupid. In Proc. 27th International Conference on Very Large Data Bases (VLDB), pages 48–58, Roma (IT), 2001.
- [23] Maedche A, Motik B, Silva N, Volz R (2002) MAFRA - A Mapping FRAMework for Distributed Ontologies. In Proc. of the 13th Intl. Conf. on Knowledge Engineering and Knowledge Management (EKAW'2002), volume 2473 of Lecture Notes in Computer Science, pages 235–250. Springer-Verlag.
- [24] Mao M, Peng Y (2006) PRIOR System: Results for OAEI 2006. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [25] Massmann S, Engmann D, Rahm E (2006) COMA++: Results for the Ontology Alignment Contest OAEI 2006. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [26] McGuinness DL, Fikes R, Rice J, Wilder S (2000) An environment for merging and testing large ontologies. Proc. 7th International Conference on Principles of Knowledge Representation and Reasoning, Breckenridge, Colorado, April 2000.

- [27] Miles A, Matthews B, Rogers N, Beckett D (2004) Skos: Standards and best practises for using knowledge organisation systems on the semantic web. In Proceedings of the NKOS Workshop: User-centred approaches to Networked Knowledge Organization Systems/Services European Conference on Digital Libraries (ECDL 2004), Bath, United Kingdom, September 2004.
- [28] Milo T, Zohar S (1998). Using schema matching to simplify heterogeneous data translation. In Proc. 24th International Conference on Very Large Data Bases (VLDB), pages 122–133, New York (NY US), 1998.
- [29] Mitra P, Wiederhold G, Kersten M (2000) A Graph-Oriented Model for Articulation of Ontology Interdependences. Proc. 7th Intl. Conf. on Extending Database Technology (EDBT 2000) Springer, Berlin-Heidelberg 2000
- [30] Niedbala S (2006) OWL-CtxMatch in the OAEI 2006 alignment contest. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [31] Noy NF (2004) Semantic Integration: A Survey Of Ontology-Based Approaches. Sigmod Record, Special Issue on Semantic Integration, 2004.
- [32] Noy NF, Musen MA (1999) SMART: Automated Support for Ontology Merging and Alignment. Twelfth Banff Workshop on Knowledge Acquisition, Modeling, and Management, Banff, Alberta, Canada, . 1999.
- [33] Noy N, Musen m (2001) Anchor-PROMPT: Using non-local context for semantic matching. In Proc. IJCAI Workshop on Ontologies and Information Sharing, pages 63–70, Seattle (WA US), 2001.
- [34] Park JY, Gennari JH, Musen MA (1998) Mappings for Reuse in Knowledge-Based Systems. 11 Workshop on Knowledge Acquisition, Modelling and Management KAW 98. Banff, Canada, 1998.
- [35] Pinto HS, Martins JP (2001) A Methodology for Ontology Integration. In Proceedings of the International Conference on Knowledge Capture K-CAP2001, pages 131-138. ACM Press, 2001.
- [36] Schorlemmer M, Kalfoglou Y (2004) A Channel-Theoretic Foundation for Ontology Coordination. In Proceedings of Meaning, Coordination and Negotiation (MCN'04) workshop of the ISWC'04, Hiroshima, Japan.
- [37] Sowa J (2000) Knowledge Representation: Logical, Philosophical and Computational Foundations. Brooks/Cole, 2000.
- [38] Studer R, Benjamins VR, Fensel D (1998) *Knowledge Engineering: Principles and Methods*. IEEE Transactions on Data and Knowledge Engineering 25(1-2):161–197
- [39] Su X (2002) A Text Categorization Perspective for Ontology Mapping. <http://www.idi.ntnu.no/~xiaomeng/paper/Position.pdf>
- [40] Zhang S, Bodenreider O (2006) NLM Anatomical Ontology Alignment System. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.
- [41] Nagy M, Vargas-Vera M, Motta E (2006) DSSim-ontology Mapping with Uncertainty. Ontology Matching workshop at International Semantic Web Conference 2006. Athens, GA, USA on November 5, 2006.